DOCUMENT RESUME

ED 177 228 TM 009 905

AUTHOR Bauer, Ernest A.; And Others
TITLE How Minimal is Minimal?
PUB DATE [Apr 79]
NOTE 15p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (63rd, San Francisco, California, April 8-12, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Standards; Complexity Level; Criterion Referenced Tests; Educational Objectives; *Equated Scores; Grade 4; Intermediate Grades; *Item Analysis; Low Ability Students; Mastery Tests; *Minimum Competency Testing; *Norm Referenced Tests; Reading Achievement; Reading Tests; State Programs; Testing Programs; *Test Interpretation; Test Items
IDENTIFIERS Comprehensive Tests of Basic Skills; Michigan; *Michigan Educational Assessment Program; *Rasch Model; Test Linking

ABSTRACT

The reading portion of the Michigan Educational Assessment Program (MEAP) was equated to the reading comprehension subtest of the Comprehensive Tests of Basic Skills (CTBS) using the Rasch Model. Both tests were administered to 366 low achieving fourth grade students. MEAP was treated as both a 95-item test and a 19-item (number of objectives attained) test. Linking constants were established by single, pairwise, and independent calibrations, as a basis for converting MEAP(95) and MEAP(19) to CTBS national percentiles. Rasch item difficulty estimates were applied to compare the tests' difficulties and to define minimal competency levels. Results indicated that the content of MEAP(19) was not reliably linked to either the CTBS or the MEAP(95), although the 19 MEAP objectives completely overlapped the CTBS. Items from tests as different in difficulty as MEAP and CTBS were successfully linked, although objective mastery was not as amenable to this procedure. (CP)

**********
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
**********

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED177228

# HOW MINIMAL IS MINIMAL?

Ernest A. Bauer
Oakland Schools

Edward J. Slawski, Jr.
Pontiac Public Schools

William R. Veitch
Oakland Schools

A paper presented at the annual convention of the
National Council on Measurement in Education,
San Francisco, 1979.

The interpretation of locally developed objective referenced tests is greatly enhanced by the fact that the tests were designed to measure the things that local educators have determined are important. If the objectives are minimal, however, interpretation becomes somewhat more difficult whether test results show that a majority of students do extremely well or extremely poorly. If only 10% of a group of students master an objective either the objective was "not really taught" or the test was "too hard." If 100% of the students master 100% of the objectives someone is likely to ask if that was because of the excellence of the students or the easiness of the test.

The interpretation of nationally normed tests has the advantage of providing a broader perspective from which to view student performance but suffers from the lack of correspondence between the content of the test and the specific content of local curriculum. Although it is meaningful for administrators to know where students in a local district stand with respect to a national sample of students, a percentile score does not tell a classroom teacher what her students know. A statewide objective referenced minimum competency assessment program suffers from the limitations of both and has the strength of neither.

At the local level, the standard approach to answering both interpretation problems has been to administer both norm referenced and objective referenced tests as part of a comprehensive testing program. Despite the increasing demands from program administrators, evaluators, curriculum designers, and the public for more testing information, teachers are resisting the intrusion of more and more tests that are only marginally (if at all) useful to them for the execution of their perceived role. The addition of a state mandated test is frequently not welcome.

One of the most promising techniques for taking advantage of the strengths of both norm referenced and objective referenced tests is available through the Rasch model. This technique enables one both to place test items on an underlying equal interval scale of content, and to equate raw scores on the tests (Rasch, 1960, 1966; Wright, 1977). The theoretical separation of the item and person parameter and the location of both parameters onto the same metric implies that substantially the same item parameters will result regardless of which students take the test and

that substantially the same achievement levels will be estimated for students regardless of which items they answer. This feature of the Rasch model allowed Wright (1968) to refer to person-free item calibration and item-free person measurement. Moreover, different tests can be 'linked' by simply determining the relationship between the origins of the scale for each test (Rentz and Bashaw, 1977). The examinees used for making this determination do not need to be a representative random sample. The model provides a methodology for determining the extent to which apparently excellent (or low) performance is due to the easiness (or difficulty) of the test or to the skills of the examinees.

The following analysis is an attempt to provide a national perspective for one of the Michigan Educational Assessment Program (MEAP) tests by using the Rasch model to link it to a nationally standardized test. It is hoped that if such a link can be successfully made between a minimum competency test and a norm referenced test, that interpretation of the objective referenced test may be facilitated both in terms of content and in terms of student performance.

## THE TESTS

The instruments used in this analysis were the reading portion of the 1977-78 fourth grade Michigan Educational Assessment Program (MEAP) and the reading comprehension subtest of the Comprehensive Tests of Basic Skills, Form S, Level 1 (CTBS). MEAP is a statewide test of 19 reading objectives (see Appendix A) which is administered to all students at the beginning of grade four. The 19 objectives represent some of the minimal skills which all fourth graders in Michigan are expected to possess. The INTERPRETATIVE MANUAL (1978) states that interpretations of the results can be made at several levels.

1. If 4 of 5 items are successfully answered, an individual student has mastered a given objective.
2. If 75% of the students in a building master 75% of the objectives, the building is a "low needs" building.
3. If, within two years, the percent of students in a building attaining 75% or more of the objectives increases by 5% and the percent of students attaining 24% or less decreases by 5%, the building is an "improving school."

While the implication exists that 100% of the students ought to master 100% of the objectives, the second and third interpretations above suggest that if an individual student attains 75% or more of the objectives, he is minimally competent. Apparently to this purpose, total number of objectives mastered by an individual student is reported. Item level data are also reported for potential diagnostic use.

The CTBS is a standardized test which measures a broad range of skills in both literal and inferential comprehension (see Appendix B). The publishers (CTB, 1971:1) state that the test measures "systematically those skills prerequisite to studying and learning in subject matter courses." The tests, moreover, are intended "for national use" and not "to measure

achievement in specific course content." In short, the tests measure "those skills common to all curricula." The vocabulary subtest was not used because the objectives tested by MEAP are predominately reading comprehension objectives. While different titles seem to refer to the same content and the same titles may refer to different content, the major content categories of the two tests appear to be quite similar.

## THE DATA

The data for the primary analysis were generated by 366 fourth grade students who attend a medium-sized suburban school district in southeastern Michigan. All had been identified as Compensatory Education (Comp Ed) students, i.e., eligible for state reimbursed compensatory education money. While Michigan law allows local districts and building considerable flexibility in the definition of Comp Ed students, buildings receive the funds based primarily on the proportion of students who attain 40% or less of the objectives on MEAP. These students, then, were expected to be low achievers in relation to the state minimal competency test. Twenty-three of these students mastered all 19 of the objectives on MEAP and were therefore eliminated for the analysis.

These students took both the CTBS and MEAP within the same month. Both tests were administered by classroom teachers following standardized administration procedures. These students also were required to take a MEAP mathematics test and the math sections of the CTBS but those tests will not be examined here.

## PROCEDURE

Proponents of the Rasch model claim that one of its most powerful features is its ability to link tests. Whenever a set of items is calibrated, the scale origin, (the mean item difficulty) is set to zero. Linking involves the determination of the difference in the scale origins of two or more tests. When one group of students takes all of the items on two tests, the tests can be calibrated as if they were one long test. The difference between the mean item difficulty of the two tests is a direct measure of the difference between the origins of the tests if calibrated independently (Rentz and Bashaw, 1977). This difference is a constant that can be added to the person ability estimates from one test to put them on the scale of the other. The reverse of that approach is to determine the mean student achievement level of the same group of students with two different tests. The difference between these means is another linking constant that can be used like the constant determined by the mean item difficulty method.

We decided to treat MEAP as if it were two tests. The test actually has 95 items (5 for each of 19 objectives) but it is almost universally treated as if it were a 19-item test. All comparative interpretations of performance typically made are based on number-of-objectives mastered, rather than any score more reflective of the 95-item nature of the test.

We will attempt to calibrate each test using BICAL (Wright and Mead, 1977) and establish linking constants for the three tests. These linking constants will be the basis for the creation of a table to convert MEAP(95) and MEAP(19) to CTBS national percentiles. More importantly, the placement of all the items on a single equal interval scale should demonstrate the extent to which the content of MEAP and the content of the CTBS overlap.

The comparison of the relative difficulty of the content of these tests will serve as an indicator of the extent to which the content of MEAP is minimal.

To validate the analysis we will compare the item difficulty estimates for the independent calibrations of each test with calibrations based on other groups of students. We will then compare the linking constant found for equating the 95-item version of MEAP [MEAP(95)] and the 19-item version of MEAP [MEAP(19)] with the link established in step 1 for Comp Ed students with the linking constant that can be determined from statewide data. This link is the only one possible to check for data at the state level without implementing a new assessment program since a standardized test is not administered as part of the statewide assessment program.

## RESULTS

### Cross-validation of the Item Difficulty Estimates

Since our primary data came from students who were very low achievers, we felt we should check our item calibrations with ones based on other students. We had access to a data set for MEAP(95) and MEAP(19) that was a statewide representative random sample (n = 5,000). We randomly selected approximately one-half of these data and calibrated each test independently. The statewide item difficulty estimates for MEAP(95) correlated .96 with the item difficulty estimates determined from the Comp Ed data; for MEAP(19) the correlation was .95.

A check for the CTBS was conducted by drawing a sample of non-Compensatory Education third and fourth grade students who had taken the CTBS. The item difficulty estimates for this group of 317 students correlated .89 with the item difficulty estimates based on the Comp Ed students. It seems that our item difficulty estimates appear to be fairly sample independent.

### Equating The Tests

Figure 1 shows the results of applying the above procedures in three ways to create the links between the three tests. The innermost triangle is based on the differences between the mean item difficulty estimates determined by a single calibration of all three tests. This approach yielded a linking constant of .97 logits for the MEAP(95) to CTBS link; and a linking constant of only .06 between MEAP(19) and CTBS.

The middle triangle in Figure 1 shows the linking constants determined by performing calibrations of pairs of tests and computing the difference between the mean item difficulty of items for each test. This approach yielded a linking constant of .94 logits for the MEAP(95) to CTBS link and .06 between MEAP(19) and the CTBS.

The outermost triangle shows the difference between mean student ability computed independently for the three tests. In each case, the three linking constants found between each pair of tests is very similar. These constants may be considered as summary measures of the distance between the content of the various sets of items as well as numbers for equating student performance.

## Is The Content Minimal?

Figure 2 is based on the pairwise calibrations shown in Figure 1. It shows the direction of the shifts in mean item difficulty as well as the degree of overlap between item difficulties of each pair of tests. It can be seen in Figure 2 that there is very little overlap in the item difficulty estimates for MEAP items and CTBS items, and MEAP items and MEAP objectives. There is almost total overlap between the MEAP objectives and the CTBS items. About one-half of both MEAP items (53%) and CTBS items (51%) are at or below the origin of their joint difficulty scale. When MEAP items are calibrated with either MEAP objectives or CTBS items, considerably less overlap results. Only 4 MEAP objectives (27%) are at or below the common origin whereas almost three quarters (73%) of MEAP items are at or below the zero point on the difficulty scale. Likewise, while four-fifths (81%) of MEAP items are easier than the average, only 9 CTBS items (25%) are at or below mean item difficulty in that joint calibration.

Figure 3 shows the distribution of item difficulties for each of the three tests when all are given a single scale origin. It can be seen in Figure 3 that only 14 items (15%) from MEAP(95) are above the common origin. Of these 14, six are among the last ten items on MEAP, which suggests that part of their difficulty may be the result of fatigue since the items are not presented in order of difficulty. Only 8 items (18%) from the CTBS are below the common origin. In other words, 85% of the items from MEAP appear to be tapping content that is at a lower level than the content tapped by 82% of the items on the CTBS. When viewed from this perspective, it is apparent that the content of MEAP(95) is quite 'minimal' in relation to the content of the CTBS.

## Is Student Performance Minimal?

Table 1 shows the results of applying the linking constants shown in Figure 1. The raw scores for given percentiles on the CTBS for the beginning of grade four were found in the publisher's manual (CTB, 1973, Pages 88-89). These raw scores were converted by using the conversion table from the independent calibration of the CTBS. In order to determine the MEAP(95) raw score that corresponds to a given CTBS raw score, the linking constant .97 was added to each CTBS log achievement estimate. These sums in logits are the MEAP(95) equivalent of the original CTBS percentile. These sums can be converted to MEAP(95) raw scores by going through the raw score to log achievement estimate table generated in the independent calibration of MEAP(95). A similar procedure was used to generate the MEAP(19) column. The linking constant used for MEAP(19) was .06.

| PER-CENTILE | NEAREST CTBS RAW SCORE | LOG ABILITY | NEAREST MEAP RAW SCORE | STATEWIDE P.R. | NEAREST NO. OBJ. MASTERED | STATEWIDE P.R. |
|---|---|---|---|---|---|---|
| 99 | 45 | | | | | |
| 95 | 42 | 2.75 | 92 | (88) | | |
| 90 | 41 | 2.44 | 91 | (83) | 17 | (58) |
| 85 | 40 | 2.19 | 90 | (78) | | |
| 80 | 39 | 1.97 | 89 | (73) | | |
| 75 | 38 | 1.79 | 88 | (67) | 16 | (49) |
| 70 | 37 | 1.63 | 87 | (63) | | |
| 65 | 36 | 1.47 | 86 | (60) | 15 | (42) |
| 60 | 34 | 1.21 | 84 | (53) | 14 | (37) |
| 55 | 32 | .97 | 82 | (47) | | |
| 50 | 30 | .75 | 79 | (40) | 13 | (33) |
| 45 | 28 | .54 | 76 | (35) | 12 | (30) |
| 40 | 25 | .25 | 72 | (30) | 11 | (27) |
| 35 | 22 | -.04 | 67 | (25) | 10 | (25) |
| 30 | 20 | -.23 | 63 | (22) | 9 | (22) |
| 25 | 18 | -.43 | 59 | (19) | 8 | (20) |
| 20 | 16 | -.63 | 55 | (17) | 7 | (18) |
| 15 | 14 | -.85 | 50 | (15) | 6 | (16) |
| 10 | 11 | -1.20 | 42 | (11) | 5 | (14) |
| 5 | 9 | -1.48 | 37 | ( 9) | 4 | (12) |
| 1 | 5 | -2.20 | 23 | ( 1) | 3 | (10) |

Table 1. The results of equating three reading tests by using the Rasch model. Entries are approximate CTBS raw score, MEAP raw score, and MEAP number-of-objectives mastered for selected percentiles of the CTBS.

We can demonstrate the use of the conversion table by examining the relationship of student scores on the various tests. Logically one can enter the table from any of several points. We will discuss two. First, the mean raw scores of the Comp Ed students on each test should convert to the same national percentile. The mean raw scores are: MEAP(19)=9; MEAP(95)=63; and CTBS=20. Inspection of the table shows that all three raw scores convert to the 30th percentile. Second, the table reveals that the state cutoff score of 75% (15) of the objectives mastered converts to the 65th percentile on the CTBS. For a population of Comp Ed students, the conversion of mean raw scores to the 30th percentile seems reasonable. These are students who, by definition, are performing well below their peers. The conversion of 75% of the objectives to the 65th percentile is more problematic.

Table 1 also shows the statewide percentile ranks for several levels of MEAP raw score and MEAP number of objectives attained. It can be seen that the relationship between MEAP(95) in terms of the Rasch equating and statewide percentile ranks is interpretable. Michigan students perform slightly better than a national standardization group, except for the bottom 15 percent. On the other hand, the only reasonable interpretation for MEAP(19) seems to be that equipercentile and Rasch equating would probably yield very different results.

If one one were to accept the MEAP(19) columns of Table 1, some questions would have to be raised about defining "minimal competencies" at a level which is higher than 65% of the national norming group. Theoretically, equating based on students for whom the tests are well targeted should yield optimum results. If equating with substantially different students yields the same results, they would take on added significance.

### Does The Table Make Sense for Statewide Data?

The availability of the statewide data and the distinction between MEAP(95) and MEAP(19) made it possible to link MEAP item scores with MEAP number-of-objectives mastered scores. Over 20% of these students had attained 19 of the 19 objectives on MEAP. These students were eliminated from the analysis. The joint calibration of MEAP(95) with MEAP(19) showed that MEAP(95) had a mean item difficulty estimate of -.11 while MEAP(19)'s was .57. The difference between these values (.68) seems substantially different from the values found for the Comp Ed data (.97; see Figure 1). However, As can be seen in Table 1, a difference of .29 (.97 minus .68) logits is never more than 5 points on the CTBS percentile scale.

The second method for linking MEAP(95) and MEAP(19) is to examine the mean student ability estimates found in independent calibrations of the two tests. The mean student ability estimate derived from MEAP(95) was 1.47; for MEAP(19) it was 1.07. These means suggest a linking constant of .40. In Figure 1, the equivalent comparison had a value of .72. These values (.40 and .72) should both be equal to the values found via the item difficulty method (.97 and .94). It is difficult to reconcile this apparent discrepancy.

A related problem for the normative interpretation of MEAP can be detected by entering Table 1 with known statewide parameters. The mean number of objectives mastered for the statewide data set is 13.7, the mean raw score is 74.9. These measures convert to scores that are 8 raw score points apart on the CTBS. Those raw score points represent a spread of 45-65 on the CTBS percentile scale. Similarly, when one enters the table with the statewide median number of objectives mastered, (16.2) one discovers that that score converts to the 75th percentile on the CTBS. This results in the interpretation that 50% of the students in Michigan would score above the 75th percentile on the CTBS, a very unlikely event.

## DISCUSSION

Our analysis of a nationally normed survey test and a state minimum competency test using the Rasch model illustrates two ways in which "minimal competency" may be examined. Using Rasch-based estimates of student achievement levels, we may equate raw scores on the separate instruments to a common equal interval scale and, thereby, define "minimal" person performance in a comparative perspective. The estimates of item difficulties which are possible using the Rasch model enables us to compare the difficulty of the content of the three instruments and to define "minimal" content. Our discussion will consider each of these applications in turn.

We demonstrated certain inconsistencies in attempts to equate

achievement scores on MEAP to a norm referenced instrument among statewide data. These inconsistencies were especially apparent using MEAP number-of-objectives mastered scores. In the case of statewide data, we found that for substantial numbers of students meaningful person achievement estimates could not be determined. However, among Comp Ed students, the relative easiness of the test was not a problem. Consequently, for applications among low achieving students (e.g., Title I evaluation), MEAP may be an appropriate test. MEAP (especially the 95-item version) could serve as an anchor test to which other norm referenced instruments could be linked, thus allowing the computation of NCE's on the basis of the administration of MEAP alone. In terms of the administration of MEAP on a statewide basis, we believe that testing efficiency could be improved by creating shorter locator tests. The authors (Slawski and Bauer, 1978) have shown that shorter tests constructed on the basis of the Rasch model can preserve most of the summary information contained in MEAP(19). Therefore, such locator tests could serve to identify those students for whom the complete test would be meaningful.

As noted above, we attempted to treat MEAP as both a 95-item test and as a 19-item (number of objectives attained) test. We were not able to reliably link the content of the 19-item test to either the CTBS or the 95-item test (even though the objective difficulties appeared to completely overlap the CTBS). Normative interpretations of the number-of-objectives mastered score are not advisable. However, while it is not difficult to understand that objective difficulties are considerably higher than the item difficulties of the same test, it seems that more work with objective referenced tests in both ways may be in order. The scale generated by the Rasch model might prove useful for establishing meaningful descriptions of what students around a cutoff score can do in terms of content. If the difficulty encountered in this analysis was due to lack of fit to the model perhaps one could alter cutoff scores to examine the effect on fit. If the problem appears to be the lack of an underlying unidimensional trait, perhaps the definitions of the objectives themselves need to be reexamined.

More encouraging in terms of the present analysis is the array of content generated from the item level data. The ability of the Rasch model to successfully scale the content of a minimal competency test and a standardized test has implications for the design of statewide assessment programs. We have demonstrated that items from tests as different in difficulty as MEAP and the CTBS can be successfully linked, while objective mastery, at least as currently defined, was not as amenable to this procedure. If a sufficient supply of calibrated items (with a substantial range of item difficulties) existed, local educational agencies could select appropriate level measures with state prescribed content. In most districts, the same content could be measured at much higher levels than that tapped by current practice. The results could be used both to identify students who need additional testing and/or remediation (if they are below a state prescribed level of achievement) and to aid teachers by providing more diagnostic information about a larger proportion of their students.

## REFERENCES

CTB/McGraw-Hill. COMPREHENSIVE TESTS OF BASIC SKILLS: EXAMINER'S MANUAL. Monterey, CA: CTB/McGraw-Hill. 1974.

INTERPRETATIVE MANUAL. Lansing, MI: Michigan Department of Education, 1978.

RASCH, G. PROBABILISTIC MODELS FOR SOME INTELLIGENCE AND ATTAINMENT TESTS. Copenhagen, Denmark: Paedogogiske Institute, 1960.

RASCH, G. An Item analysis which takes individual differences into account. BRITISH JOURNAL OF MATHEMATICAL AND STATISTICAL PSYCHOLOGY, 1966, 19(1), 49-57.

RENTZ, R. and BASHAW, W. "The national reference scale for reading: An application of the Rasch model", JOURNAL OF EDUCATIONAL MEASUREMENT, 1977, 14(2), 161-180.

SLAWSKI, E. and BAUER, E. "Reducing testing time without loss of test information: A ten-item MEAP test." Paper presented at the Michigan Educational Research Association Annual Meeting, Detroit, MI, 1978.

WRIGHT, B. "Sample-free test calibration and person measurement." In PROCCDINGS OF 1967 INVITATIONAL CONFERENCE ON TESTING PROBLEMS. Princeton, NJ: ETS, 1968, 85-101.

_____. "Solving measurement problems with the Rasch Model." JOURNAL OF EDUCATIONAL MEASUREMENT, 1977, 14(2), 97-116.

_____ and MEAD, R. BICAL: RESEARCH MEMORANDUM NO. 23. Chicago: Statistical Laboratory, Department of Education, University of Chicago.

## APPENDIX A

## MINIMAL PERFORMANCE READING OBJECTIVES MEASURED IN THE 1977-78 MICHIGAN EDUCATIONAL ASSESSMENT PROGRAM GRADE FOUR

| | | |
|---|---|---|
| Vocabulary Meaning | 15 items | (16%) |

Match words with definitions
Indicate phrases with same meanings
Choose word appropriate to blank spaces (cloze procedure)

| | | |
|---|---|---|
| Literal Comprehension | 30 items | (32%) |

Indicate factual selections
Indicate fictional selections
Indicate title most appropriate for selection
Indicate pictures best describing main idea in selections
Choose phrase best describing work in figurative language
Match causes with effects

| | | |
|---|---|---|
| Inferential Comprehension | 30 items | (32%) |

Indicate author's purpose
Choose best summary of a selection
Match quotation from story with speaker
Choose answer best describing how character feels in story
Choose most appropriate conclusion for a story
Select meaning, generalizations, conclusions not expressed

| | | |
|---|---|---|
| Study Skills | 20 items | (21%) |

Identify method of arranging data
Alphabetize words through first 3 letters
Answer locational question about reference sources
Answer locational question about newspapers

## APPENDIX B

## CONTENT CATEGORIES FOR CTBS READING COMPREHENSION, LEVEL 1

Literal Recall 12 items (27%)

Recall facts and details explicitly stated in a passage. The student is required to answer who, what, where, when questions.

Rewording 8 items (18%)

Given facts and details in a passage, answer questions by choosing paraphrased or reworded material.

Context Clues 2 items (4%)

Determine word meaning from the context in which a word is used.

Main Idea 8 items (18%)

Identify the topic of a passage, restate or summarize the central thought in a passage, identify the author's purpose in writing the passage, derive a lesson or moral from the passage, or select the best title for a passage.

Descriptive Words 4 items (9%)

Define the physical attributes of an object or person presented in the passage, analyze emotions experienced in the passage, or decide what quality best characterizes a person.

Conclusions 11 items (24%)

Draw conclusions based on facts stated or implied in the passage or perceive cause-effect relationships between events or ideas.

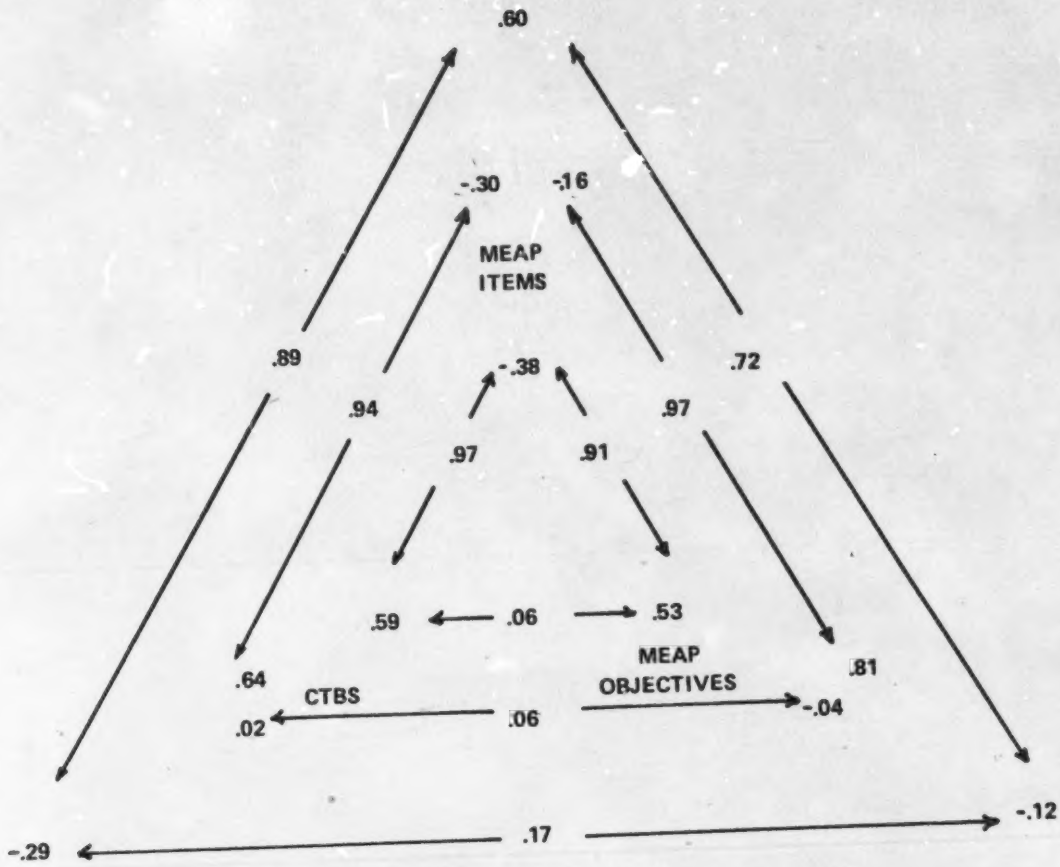


Figure 1. The results of different methods of determining the linking constants between three tests. The inner triangle is based on the mean item difficulties from a single calibration of all three tests. The middle triangle is based on the mean item difficulties from three pairwise calibratrions. The outer triangle is based on the mean student ability estimate from each of three independent calibrations.

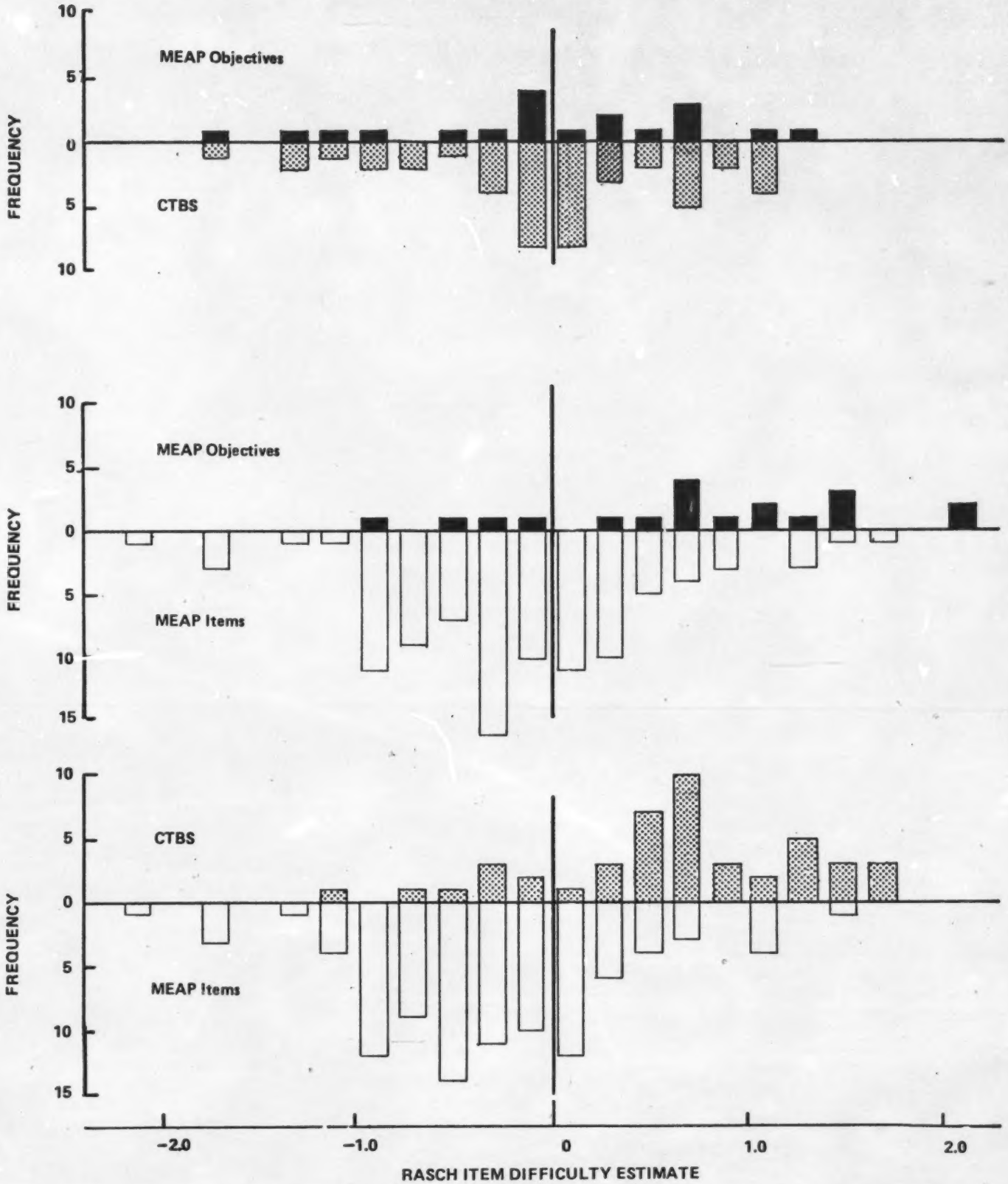


Figure 2. The distributions of MEAP items, MEAP objectives and CTBS Reading Comprehensions pairwise RASCH item calibrations.

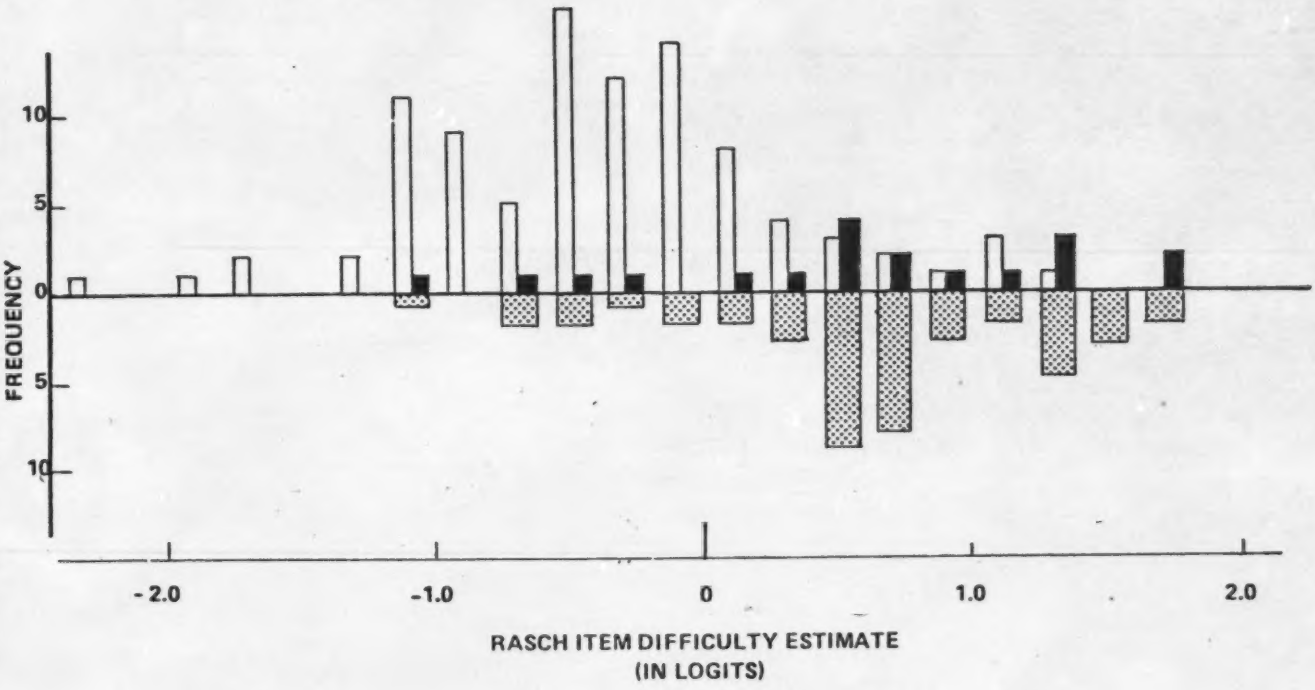


**Figure 3.** The distribution of MEAP items, MEAP objectives, and CTBS Reading Comprehension item difficulty estimates from a single calibration of all three tests.

■ = MEAP Objectives
□ = MEAP Items
▨ = CTBS